Simplifying Low-Light Image Enhancement Networks with Relative Loss Functions

Yu Zhang[1], Xiaoguang Di[1*], Junde Wu[2], Yue Wang[1], Yong Li[1], Rao Fu[3], Yanwu Xu[2], Guohui Yang[1] and Chunhui Wang[1*]

[1*]Harbin Institute of Technology, Harbin, 150001, China.
[2]South China University of Technology, Guangzhou, 510000, China.
[3]Mind Vogue Lab, Beijing, 100000, China.

*Corresponding author(s). E-mail(s): dixiaoguang@hit.edu.cn; wang2352@hit.edu.cn;
Contributing authors: zhangyuhit2@hit.edu.cn; jundewu@ieee.org; seraph@hit.edu.cn;
522384058@qq.com; 1098712828@qq.com; ywxu@ieee.org; gh.yang@hit.edu.cn;

Abstract

Image enhancement is a common technique used to mitigate issues such as severe noise, low brightness, low contrast, and color deviation in low-light images. However, providing an optimal high-light image as a reference for low-light image enhancement tasks is impossible, which makes the learning process more difficult than other image processing tasks. As a result, although several low-light image enhancement methods have been proposed, most of them are either too complex or insufficient in addressing all the issues in low-light images. In this paper, to make the learning easier in low-light image enhancement, we introduce FLW-Net (Fast and LightWeight Network) and two relative loss functions. Specifically, we first recognize the challenges of the need for a large receptive field to obtain global contrast and the lack of an absolute reference, which limits the simplification of network structures in this task. Then, we propose an efficient global feature information extraction component and two loss functions based on relative information to overcome these challenges. Finally, we conducted comparative experiments to demonstrate the effectiveness of the proposed method, and the results confirm that the proposed method can significantly reduce the complexity of supervised low-light image enhancement networks while improving processing effect. The code is available at https://github.com/hitzhangyu/FLW-Net.

Keywords: Low-light Image, Image Enhancement, Lightweight Network, Relative Loss.

1 Introduction

Images captured in dark environments or with insufficient exposure often become low-light images that suffer from low contrast, low brightness, severe noise, and color deviation, making some information in the images invisible.

To improve the quality of these images, numerous low-light image enhancement methods have been proposed in recent years (C. Guo et al., 2020; Y. Jiang et al., 2021; X. Liu, Xie, Zhao, Wang, & Meng, 2023; Ma, Ma, Liu, Fan, & Luo, 2022; Y. Zhang, Guo, Ma, Liu, & Zhang, 2021). Although these methods have shown promising

results, there remains a trade-off between processing speed and effect, regardless of whether they are based on learning or not.

While non-learning-based low-light image enhancement methods are capable of significantly improving image contrast and brightness(X. Guo, Li, & Ling, 2016; Kimmel, Elad, Shaked, Keshet, & Sobel, 2003; Lee, Lee, & Kim, 2013; Yu & Zhu, 2019), subsequent separate denoising steps (Dabov, Foi, Katkovnik, & Egiazarian, 2007) or joint iterative denoising processes based on variation (M. Li, Liu, Yang, Sun, & Guo, 2018) can be time-consuming. This limitation makes most of these methods unsuitable for real-time low-light image enhancement applications.

Learning-based low-light image enhancement methods can be categorized into two main groups: the supervised learning-based and the unsupervised learning-based. Typically, unsupervised learning-based methods are lightweight and are more robust to various environments (C. Guo et al., 2020; C. Li, Guo, & Loy, 2021; Ma et al., 2022). However, they often lack responsive color correction and denoising methods, which limits their ability to improve the accuracy of subsequent high-level tasks (Y. Zhang, Di, Zhang, Ji, & Wang, 2022). In contrast, supervised learning-based methods are designed to address all types of low-light image degradation and can significantly improve image quality (N. Jiang, Lin, Zhang, Zheng, & Zhao, 2023; K. Zhang, Yuan, Li, Gao, & Li, 2023; Zhou, Shi, & Ren, 2023). However, most of these methods require a complex network structure, resulting in longer processing times.

Designing a lightweight network that can simultaneously enhance contrast and remove noise is a challenging task. Two primary challenges limit the simplification of the network. Firstly, the contrast adjustment of an image should consider both local and global information. This requires a large receptive field to capture global contrast, which increases the complexity of the network. Secondly, in low-light image enhancement, there is no absolute reference, which means the network must learn uncertain output values for the same input pixel or image block during denoising (We can name it the **one-to-many problem**). That makes it more difficult for supervised learning-based methods to learn compared to unsupervised learning-based methods, since most unsupervised learning-based methods have a consistent assumption about the output value, such as the average brightness value being close to a fixed value(C. Li et al., 2021).
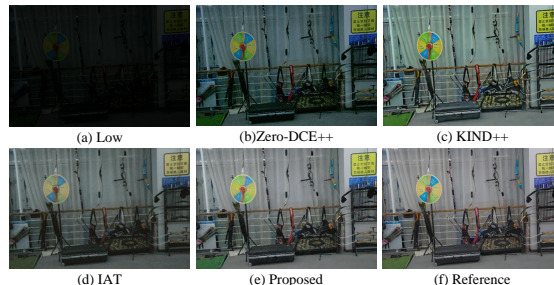

Fig. 1 Visual comparison with some SOTA methods. (a) The original input low-light image. (b) to (e) are enhanced images produced by Zero-DCE++ (C. Li et al., 2021), KIND++ (Y. Zhang, Guo, et al., 2021) and the proposed method in this paper. (f) The reference image. It can be seen that the color of the enhanced image (e) produced by our proposed method is the closest to that of the reference image (f).

In this paper, to address these challenges, we propose a solution by introducing a fast and lightweight image enhancement network called FLW-Net, along with two specially designed loss functions based on relative information for the low-light image enhancement task. We can abbreviate the losses based on relative information as the relative losses, and relative losses can achieve an extremum when the reference and the output are similar in some features, rather than requiring them to be exactly equal. An example of enhancing a low-light image comprising color deviation is shown in Fig. 1. It can be seen that the color of the enhanced image produced by proposed method is the closest to that of the reference image.

Specifically, in terms of network structure design, we propose a simple Global Feature Extraction (GFE) component that can extract global information from the image's histogram and generate a global brightness adjustment proposal through a higher-order curve adjustment method (C. Guo et al., 2020). In terms of loss functions, we propose two losses based on relative information to alleviate the one-to-many problem in low-light image enhancement task.

Our contributions can be summarized as follows:

- To improve computational efficiency by utilizing fewer parameters for obtaining global information, we proposed a Global Feature Extraction (GFE) component with only 1.4K parameters. Unlike most existing methods that use large receptive fields or Transformer structures for extracting global information, GFE extracts information from histograms with a small number of bins. This not only makes it more efficient but also allows for easier integration with hyperparameters that control the degree of global enhancement.
- To address the learning difficulty caused by multiple potential reference images, we proposed two novel loss functions ($L_{brightness}$ and $L_{structure}$) based on relative information. Unlike traditional L_1 or L_2 losses that aim to make the enhanced image exactly match the reference image, our proposed loss functions utilize cosine similarity, allowing for brightness differences between the output and reference. Then the network can focus more on color, structure and noise removal. When the proposed loss functions are combined with other supervised methods, it can achieve better quantitative metrics with less parameters(e.g., only about 5% parameters of the original KIND's network when combined with KIND (Y. Zhang, Zhang, & Guo, 2019)).
- With GFE and two proposed loss functions, we built a fast and lightweight low-light image enhancement network, named FLW-Net, which achieves comparable or even better performance compared to state-of-the-art methods while maintaining faster processing speed. This shows the potential of reducing the complexity of low-light image enhancement networks with the proposed methods.

2 Related Work

2.1 Low-Light Image Enhancement

Non-learning-based low-light image enhancement solutions mainly include histogram equalization, gamma correction, methods based on dehazing (Dong et al., 2011) or the Retinex model (S. Wang, Zheng, Hu, & Li, 2013), as well as other improved methods based on these approaches (Celik & Tjahjadi, 2011; G. Fu, Duan, & Xiao, 2019; X. Guo et al., 2016; Kumar & Bhandari, 2022; Park, Yu, Moon, Ko, & Paik, 2017).

Although these methods can significantly improve the brightness and contrast of images, removing noise and restoring color remain challenging.

Recently, learning-based methods for enhancing low-light images have achieved promising results, including supervised methods (Xu, Chen, Xu, Jin, & Zhu, 2022; Y. Zhang et al., 2022; Y. Zhang, Guo, et al., 2021; Y. Zhang et al., 2019) and unsupervised methods (C. Guo et al., 2020; Y. Jiang et al., 2021; X. Liu, Ma, Ma, & Wang, 2023; Ma et al., 2022; Xiong, Liu, Shen, Fang, & Luo, 2020; Y. Zhang, Di, et al., 2021). However, unlike other image processing or computer vision tasks, the low-light image enhancement task usually lacks an ground-truth/absolute label. For the same scene, there may be multiple low-light and high-light images, making it difficult to determine which reference image is the best. Even after expert correction, it can still be challenging to select the ideal reference image (Y. Zhang et al., 2022; Y. Zhang, Guo, et al., 2021).

Therefore, most supervised methods for low-light image enhancement often face challenges due to the presence of multiple potential reference images. To address this issue, there are mainly two types of methods. The first involves designing complex networks, such as the current state-of-the-art method MAXIM (Tu et al., 2022). While these methods can achieve promising visual results, they are often time-consuming for low-light image enhancement tasks.

The second type of method involves connecting the input and output during training by introducing hyperparameters (Chen, Chen, Xu, & Koltun, 2018; Q. Fu, Di, & Zhang, 2020), simplified Retinex models (Wei, Wang, Yang, & Liu, 2018), or both (Y. Zhang et al., 2019). For instance, Chen et al. (2018) introduced the exposure time ratio of reference and input images as a hyperparameter to achieve denoising and color restoration. Q. Fu et al. (2020) further proposed a sub-network to automatically select hyperparameters. Wei et al. (2018) incorporated the simplified Retinex model into the network. However, the assumption in the simplified Retinex model that all three color channels have the same illumination image does not align with reality (Y. Zhang et al., 2022), leading to suboptimal denoising effects.

To address this issue, Y. Zhang et al. (2019) and Y. Zhang, Guo, et al. (2021) introduced both hyperparameters and the simplified Retinex model

into the network and designed a separate restoration module to remove noise and correct color in the reflection image. However, it is important to note that the introduction of both Retinex models and hyperparameters might still deviate from the real imaging process. Consequently, while these methods involve less complex network structures, their running time may not meet the requirements of practical applications.

Unsupervised methods for low-light image enhancement typically assume that the output meets certain constraints, making them more lightweight and more stable for unseen scenes. For example, C. Guo et al. (2020) use the mean value assumption (e.g., supposing the mean brightness of the image is between 0.4 and 0.6) and some specially designed loss functions to constrain the output. Xiong et al. (2020) specify the initial value of the illumination image in the simplified Retinex model as the max value of R,G,B in each pixel. Y. Jiang et al. (2021) propose to learn constraints on the output from the normal-light images through GAN framework. Ma et al. (2022) proposed to constrain the similarity of outputs at different stages during training. Although most of these unsupervised methods can meet real-time requirements, accurately removing noise and restoring color remains a challenge due to the lack of sufficient noise and color constraints.

By comparing supervised and unsupervised methods, it can be observed that if we do not aim to achieve enhancement results that are identical to the reference image but rather focus on learning color, structure, and noise removal from the reference image, it is possible to utilize a simple network to effectively remove noise and restore colors during image enhancement. At this point, the problem to be addressed is how to design evaluation metrics or loss functions that are not affected by brightness differences.

2.2 Image Retouching

Image retouching methods focus on problems such as inappropriate brightness, poor contrast, color deviation, etc., similar to image enhancement tasks (Y. Wang, Li, et al., 2022). However, most of these methods do not consider the noise problem. Therefore, basic retouching operations can work on a single pixel, making them extremely fast and lightweight (He, Liu, Qiao, & Dong, 2020;

Y. Liu et al., 2022; Y. Wang, Li, et al., 2022; Zeng, Cai, Li, Cao, & Zhang, 2020). For example, He et al. (2020) and Y. Liu et al. (2022) proposed the Conditional Sequential Retouching Network (CSRNet) with only 37K trainable parameters. Y. Wang, Li, et al. (2022) proposed the trainable neural color operators, which contains only 28K parameters in their method. The successful applications of simple networks in image retouching also demonstrate the feasibility of using simple networks for image enhancement tasks.

3 Methodology

Figure 2 illustrates the architecture of FLW-Net, which comprises two primary modules: the Global Feature Extraction (GFE) component and the Local Enhancement Network (LEN) component. The GFE component takes the low-light image's V channel and the desired average brightness as inputs and produces a global brightness adjustment proposal through higher-order curve adjustment method. Then, the proposal is concatenated and fed into LEN. The LEN component takes the low-light image and the global brightness adjustment proposal as inputs and enhances the image with some carefully designed loss functions. It consists of several convolutional layers with a local receptive field to capture local information and generate high-frequency details.

The proposed method includes several loss functions beyond the commonly used L_1 and SSIM. The color loss, denoted as L_{color}, is used to measure the color similarity between the enhanced image and the reference image. The brightness loss, denoted as $L_{brightness}$, is used to measure the difference in brightness orders between the enhanced image and the reference image. Finally, the structure loss, denoted as $L_{structure}$, is designed to encourage the enhanced image to have similar gradient orders to the reference image. We refer to these losses as relative losses.

Unlike in L_1 and SSIM losses, where the extremum is only reached when the output and reference are exactly equal, the relative loss can achieve the extremum when the output and reference share some common features. All of these loss functions, including the L_1 and SSIM loss, are combined to form the total loss function used in training the FLW-Net. To the best of our knowledge, $L_{brightness}$ and $L_{structure}$ are first proposed
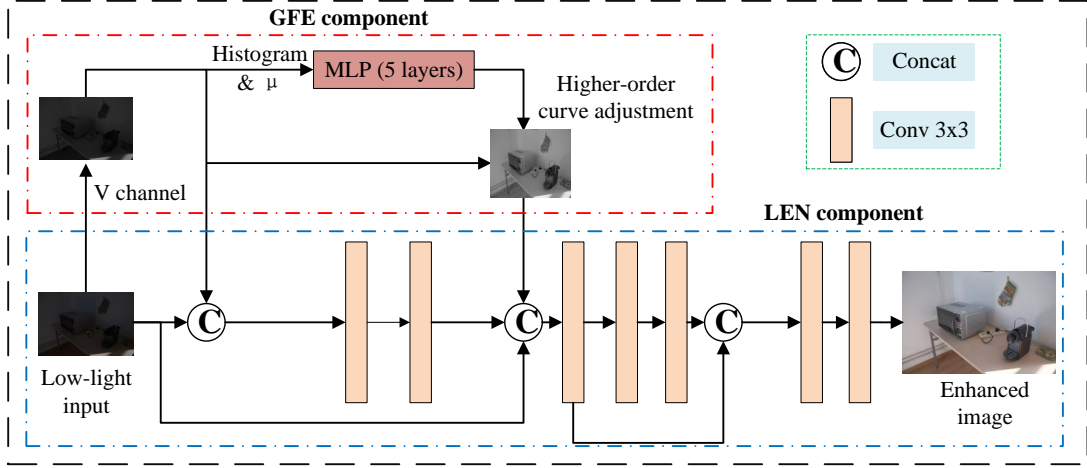
Fig. 2 The detailed structure of the proposed method.

by this paper in this low-light image enhancement task.

3.1 Global Feature Extraction Component

It is unnecessary to emphasize the importance of global information extraction in low-light image enhancement, as it has been extensively discussed in previous literature (Cui et al., 2022). However, the challenge lies in efficiently extracting global information and integrating it into the enhancement network.

Y. Zhang et al. (2022) has proven that the V channel in the HSV color space is sufficient to represent the brightness of the input low-light image. Meanwhile, C. Guo et al. (2020) proposed to iteratively apply Equation (1) to adjust the input low-light image.

$$\mathbf{I_{k+1}} = \mathbf{I_k} + \alpha_k \mathbf{I_k}(1 - \mathbf{I_k}) \tag{1}$$

where k represents the number of iterations (e.g., $\mathbf{I_0}$ represents the input low-light image).

Inspired by those two works, we propose to extract global information from the histogram of the V channel of the low-light image and represent it as higher-order curve coefficients, denoted as $\alpha_{0,1,\dots t}$. Specifically,

$$\{\alpha_{0,1,\dots t}\} = G(H(\mathbf{I^v})) \tag{2}$$

where $\mathbf{I^v}$ represents the V channel of the low-light image $\mathbf{I}$ in the HSV color space, $H(\mathbf{I^v})$ represents the operation of obtaining the histogram of the image $\mathbf{I^v}$, and $G(\cdot)$ represents the Global Feature Extraction (GFE) component, which is implemented using a Five-layer Multi-Layer Perception with only 1.4K parameters. Compared with the method of C. Guo et al. (2020), the GFE component in our proposed method extracts global information from the histogram of the V channel instead of at the pixel level, which is more efficient. Specifically, the computational cost of the GFE component does not significantly increase with the increase in image size. Additionally, the output of GFE component is not the final enhanced result, which allows for further denoising and color restoration. Compared with the method of Y. Zhang et al. (2022), the proposed method does not need to be combined with other image contrast enhancement methods and can achieve end-to-end training with the help of the loss functions proposed in Eqn. (9) and Eqn. (11).

As mentioned, there is currently no definitive standard that defines whether an enhancement result is optimal. Similarly, we cannot claim that the processing result of a fixed-parameter network trained on a specific dataset is optimal. In this case, introducing adjustable parameters for image enhancement is a reasonable solution, as adopted in previous work (X. Liu, Xie, et al., 2023; Y. Wang, Wan, et al., 2022; Y. Zhang, Guo, et al.,

2021). Additionally, if the parameters are related to reference images, we can allow the network to learn relatively deterministic mapping operations, thereby simplifying the training difficulty of the network in the training phase.

The GFE component can be easily modified to incorporate additional hyperparameters. In this paper, we used the desired average brightness value, μ, as the hyperparameter. During the training process, μ is calculated as the mean value of the V channel of the reference image. Compared to previous works which utilized parameters such as exposure time difference between reference images and input (Chen et al., 2018), and brightness ratio between the illumination image of the reference image and that of the input image (Y. Zhang et al., 2019), μ is more intuitive and highly correlated with histogram information. Moreover, during the testing process, we can easily fix μ to a constant value, as done in some unsupervised works(C. Guo et al., 2020). Therefore, the modified GFE component can be expressed as:

$$\{\alpha_{0,1,...t}\} = G(H(\mathbf{I^v}), \mu) \qquad (3)$$

Then, we can iteratively adjust the V channel of the low-light image to obtain the global brightness adjustment proposal, as follows:

$$\mathbf{I^v_{k+1}} = \mathbf{I^v_k} + \alpha_k \mathbf{I^v_k}(1 - \mathbf{I^v_k}) \qquad (4)$$

After obtaining the global brightness adjustment proposal, it will be concatenated with the middle layer of LEN and fed into the LEN for further enhancement. The entire FLW-Net is trained end-to-end, which means that all the components are trained jointly to optimize the overall performance of the network.

3.2 Loss Functions based on Relative Information

Let us consider $\{\mathbf{I}, \mathbf{Y}\}$ as one paired low/high-light images. Typically, we use lots of paired images to train the enhancement network E and hope it can well fit the following Equation (5):

$$\mathbf{Y_{(i,j)}} = E(\mathbf{I_{(i,j)}}) \qquad (5)$$

where (i, j) represents the coordinate of one pixel. To achieve this, various loss functions have been adopted to train the enhancement network E with paired low/high-light images, including commonly used L_1, L_2 and SSIM loss, etc. However, due to the one-to-many problem, these loss functions may not be as effective for image enhancement as they are for other low-level image processing tasks such as image denoising, image deblurring, and image dehazing. In other words, these loss functions are more suitable for learning the mapping relationship with an absolute reference image for the input.

As previously mentioned, previous studies have proposed strategies to establish a one-to-one relationship between the input and output of the enhancement models (where one input low-light image corresponds to one certain reference image). However, the Retinex models and additional parameters adopted in those work cannot onvert the one-to-many problem into a real one-to-one problem. In practice, even adding a hyperparameter for each pixel does not guarantee a satisfactory color restoration effect(Y. Zhang et al., 2022). Therefore, most of previous works often require complex networks or separate denoising modules, such as SID (Chen et al., 2018), RetinexNet (Wei et al., 2018), KinD (Y. Zhang et al., 2019), KinD++ (Y. Zhang, Guo, et al., 2021).

Since there is no absolute supervision information, one intuitive approach is to use relative information in the loss functions, which reduces the assumption of existence of absolute reference images. Previous unsupervised methods have proposed some useful loss functions, such as spatial consistency loss (C. Guo et al., 2020), normalized gradient error (Y. Zhang, Di, et al., 2021), and perception loss (Y. Jiang et al., 2021). However, most of them do not achieve impressive results in supervised methods, since they impose weak constraints.

In this paper, we propose the use of relative losses, which are loss functions based on relative information, to make the learning easier for networks.

Firstly, for one pixel, we expect its color information more than the brightness to match the reference image. To achieve this, we need to extract the image's color information, which can be accomplished through various color spaces, such as the HSV (Hue, Saturation, Value) or HSI (Hue, Saturation, Intensity) color space. In this paper, we chose to use the HSV color space. It

has been demonstrated that two pixels share the same Hue and Saturation when they satisfy the following equation (Y. Zhang et al., 2022):

$$E(\mathbf{I_{(i,j)}}) = \lambda \mathbf{Y_{(i,j)}} \tag{6}$$

where λ represents arbitrary non-zero positive number. It should be noticed that both $E(\mathbf{I_{(i,j)}})$ and $\mathbf{Y_{(i,j)}}$ represent 3D vectors. Then, we can adopt the cosine similarity to measure the Hue and Saturation difference between two pixels. Therefore, the loss function for color can be designed as follows:

$$L_{color} = 1 - \sum_{i=1,j=1}^{m,n} < E(\mathbf{I_{(i,j)}}), \mathbf{Y_{(i,j)}} > \tag{7}$$

where $< \cdot, \cdot >$ represents the cosine similarity of two vectors. By minimizing this loss function, the network is encouraged to match the Hue and Saturation information between the output and reference images. This loss function has also been adopted in many other image enhancement (C. Liu, Wu, & Wang, 2022; K. Zhang et al., 2023) and Computational Color Constancy work (Barnard, Cardei, & Funt, 2002; Gijsenij, Gevers, & van de Weijer, 2011).

Secondly, regarding the brightness, it is expected that the enhanced images have the same lightness order as the reference (S. Wang et al., 2013), which means that images that are brighter in the reference should also be brighter in the enhanced image. S. Wang et al. (2013) proposed the evaluation metric, LOE (Lightness-Order-Error), for the lightness order error. However, directly incorporating LOE into the loss function is not straightforward since it is difficult to calculate the corresponding gradients in back propagation. In this paper, we propose the following equation to model the brightness relation between the enhanced image and the reference.

$$b(E(\mathbf{I_{(i,j)}})) = \beta b(\mathbf{Y_{(i,j)}}) + \gamma \tag{8}$$

where $b(\cdot)$ represent image blocks centered on pixels $E(\mathbf{I_{(i,j)}})$ and $\mathbf{Y_{(i,j)}}$, β and γ can represent 3D vectors for color image or scalars in gray images. For different blocks, there can be different β and γ. This is useful for processing images with non-uniform brightness, where different regions of the image have different enhancement levels. It can be

seen that, in this model, the enhanced images have the same lightness order as the reference image in every image block. Also, it is more rigorous than the LOE metric since it requires more than just brightness order. Specifically, it requires the enhanced image to be a linear transformation of the noise-free reference image, which significantly suppresses noise. Then, we can design the loss function as follows:

$$L_{brightness} = 1 - \sum_{c \in R,G,B} \sum_{i=1,j=1}^{m,n}$$
$$< b(E(\mathbf{I_{(i,j)}^c})) - min(b(E(\mathbf{I_{(i,j)}^c}))),$$
$$b(\mathbf{Y_{(i,j)}^c}) - min(b(\mathbf{Y_{(i,j)}^c})) > \tag{9}$$

where c represents the different color channels, and the purpose of subtracting the minimum value is to remove the influence of the constant γ. The following experimental results show that $L_{brightness}$ can improve the PSNR metric which is related to noise suppression(In Sec. 4.3, Table 4 and Figure 9).

Thirdly, for the structure information, it is usually expressed by gradient information. We can adopt a similar model as Equation (8). The difference is replacing the brightness value with gradient. Then, we can get the following Equation (10).

$$b(\nabla E(\mathbf{I_{(i,j)}})) = \eta b(\nabla \mathbf{Y_{(i,j)}}) + \epsilon \tag{10}$$

Then, the loss funciton for structure can be expressed as follows:

$$L_{structure} = 1 - \sum_{c \in R,G,B} \sum_{i=1,j=1}^{m,n}$$
$$< b(\nabla E(\mathbf{I_{(i,j)}^c})) - min(b(\nabla E(\mathbf{I_{(i,j)}^c}))),$$
$$b(\nabla \mathbf{Y_{(i,j)}^c}) - min(b(\nabla \mathbf{Y_{(i,j)}^c})) > \tag{11}$$

If we remove the ϵ part in Eqn.(10), Eqn.(10) will become the derivative of the Eqn.(8). At this point, $L_{brightness}$ and $L_{structure}$ will essentially represent the same thing. However, by introducing ϵ in Eqn.(10), it will allow for non-linear changes in brightness within image blocks. This more closely aligns with the imaging process in reality and can reduce the effects of saturation and under-saturation caused by linear changes in brightness during enhancement. As a result, $L_{structure}$ is more effective in preserving image
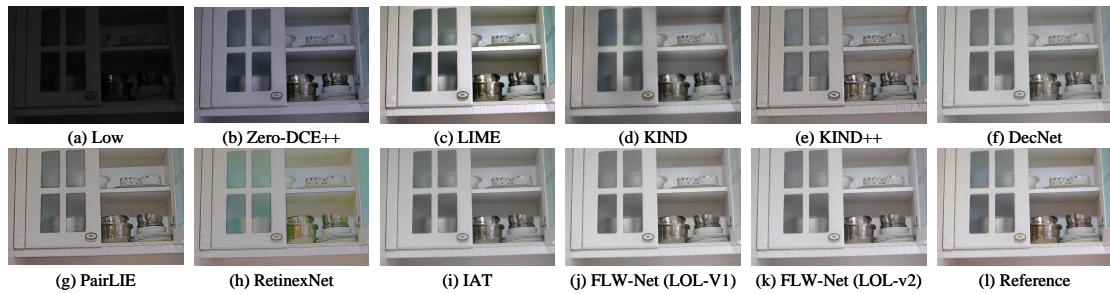
Fig. 3 Visual comparison results on LOL-V1 dataset. Comparison methods include Zero-DCE++, LIME, KIND, KIND++, DecNet, PairLIE, RetienxNet, IAT and the proposed FLW-Net trained on LOL-V1 and LOL-V2 datasets. (Best viewed on high-resolution displays with zoom-in.)

structures. Experimental results also indicate that it is more effective in improving the structural similarity index (SSIM) metric(In Sec. 4.3, Table 4 and Figure 9).

The total loss can be expressed as follows:

$$L_{ALL} = L_1 + L_{SSIM} + L_{color} + L_{brightness} + L_{structure}$$
(12)

where L_{SSIM} represents the SSIM loss between the enhanced and reference images.

4 Experiments

4.1 Implementation Details

The framework is implemented with PyTorch on an NVIDIA 3090 Ti GPU. The batch size used for training is 171. We use the Adam optimizer to train the network with a learning rate of 0.0001. We mainly use two datasets for training and testing: the LOL-V1 dataset (Wei et al., 2018) and the LOL-V2 dataset (Yang, Wang, Huang, Wang, & Liu, 2021). The LOL-V1 dataset contains 500 image pairs, with 15 pairs used for testing. The LOL-V2 dataset contains 689 image pairs for training and 100 pairs for testing. We also collected some images online and merged them together, which we refer to as the Mixed dataset. The Mixed dataset includes the test images of LOL-V1 (15 images), LIME (X. Guo et al., 2016) (10 images), MF (X. Fu et al., 2016) (10 images), and VV[1] (23 images), and most of these images do not have corresponding reference images.

We used four metrics for quantitative comparison: PSNR, SSIM, CIEDE2000 (Luo, Cui, & Rigg,

2001; Sharma, Wu, & Dalal, 2005), and NIQE (Mittal, Soundararajan, & Bovik, 2012). PSNR and SSIM are reference image quality assessment methods that indicate the noise level and the structural similarity between the enhanced images and the reference, respectively. CIEDE2000 is a reference image quality assessment method to accurately measure colors differences, which is published by International Commission on Illumination(Also known as the Commission Internationale de l´Eclairage, CIE), and a lower value indicates less color difference. NIQE is a nonreference image quality assessment method that evaluates the naturalness of the image, and a lower value indicates better quality. To differentiate between the μ values during the training and testing processes, we use μ_{train} and μ_{test} to represent them, respectively.

4.2 Objective Performance Evaluation

In this section, we compared our method with several state-of-the-art (SOTA) low-light image enhancement methods, including LIME (X. Guo et al., 2016), RetinexNet (Wei et al., 2018), Zerodce++ (C. Li et al., 2021), KIND (Y. Zhang et al., 2019), KIND++ (Y. Zhang, Guo, et al., 2021), DecNet (X. Liu, Xie, et al., 2023), PairLIE (Z. Fu et al., 2023) and IAT (Cui et al., 2022). Among them, LIME is a non-learning-based method, and ZeroDCE++ can be trained without any references. PairLIE can be trained with paired low-ligh images. The other methods are based on supervised learning. Among them, KIND (Y. Zhang et al., 2019), KIND++(Y. Zhang, Guo, et al., 2021) and DecNet(X. Liu, Xie, et al., 2023)

---

[1]https://sites.google.com/site/vonikakis/datasets/challenging-dataset-for-enhancement

Table 1 Quantitative comparison results on LOL (V1 (Wei et al., 2018) & V2 (Yang et al., 2021)) datasets and a mixed dataset. The mixed dataset includes the test images of LOL-V1 (15 images), LIME (X. Guo et al., 2016) (10 images), MF (X. Fu et al., 2016) (10 images), and VV (23 images), and most of these images do not have corresponding reference images. It should be noted that the test time of KIND and KIND++ comes from KIND++(Y. Zhang, Guo, et al., 2021) with a Titan-X GPU, and the test time of RetinexNet does not include the running time of BM3D(Dabov et al., 2007) for denoising.

Method	LOL-V1			LOL-V2			Mixed Dataset(Unpaired)	Efficiency	
	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	NIQE↓	Params(M)↓	test time(s)↓
LIME(X. Guo et al., 2016)	17.22	0.50	5.32	15.77	0.46	5.37	4.57	-	0.190
RetinexNet(Wei et al., 2018)	17.86	0.78	6.37	17.37	0.76	9.09	5.68	0.4	0.019*
Zerodce++(C. Li et al., 2021)	15.35	0.57	7.86	18.49	0.58	8.05	4.53	**0.01**	**0.001**
KIND(Y. Zhang et al., 2019)	20.38	0.83	5.45	23.78	**0.88**	4.96	3.87	8.21	0.11*
KIND++(Y. Zhang, Guo, et al., 2021)	21.80	**0.84**	5.17	22.21	0.84	4.89	3.74	8.28	0.12*
IAT(Cui et al., 2022)	23.38	0.81	**3.92**	23.50	0.82	4.29	4.71	0.09	0.004
DecNet(X. Liu, Xie, et al., 2023)	22.49	0.82	4.51	22.56	0.84	4.83	4.26	1.83	0.353
PairLIE(Z. Fu et al., 2023)	18.47	0.75	4.25	19.88	0.78	4.34	3.90	0.33	0.057
FLW(Training on LOL V1)	23.84	0.83	4.22	25.71	0.87	4.09	3.93	0.02	0.001
FLW(Training on LOL V2)	**24.70**	**0.84**	4.11	**26.61**	**0.88**	**3.89**	**3.72**	0.02	**0.001**

all have hyperparameters during training and testing. The comparison results are shown in Table 1, 2 and 3 and Figures 3, 4, 5 and 6. In Fig. 5 and 6, the hyperparameter μ is fixed at a constant value of 0.4.

During the training on the LOL-V1 dataset, we utilized only 343 images, which is approximately half of the LOL-V2 training data. It should be noted that both datasets were produced by the same team, and LOL-V2 contains most of the data in LOL-V1. Hence, we can evaluate the impact of training data volume on the network.

As shown in Table 1, the training data volume has a greater impact on PSNR than SSIM. For instance, when trained on the LOL-V1 dataset, the PSNR reduced by almost 0.9 dB compared to the LOL-V2 dataset (PSNR: 26.61 → 25.71). However, the SSIM only decreased by 0.01 (SSIM: 0.88 → 0.87). Furthermore, irrespective of whether trained on the LOL-V1 or LOL-V2 dataset, FLW-Net outperforms other methods in terms of PSNR. Regarding SSIM, FLW-Net, KIND, and KIND++ achieve similar results. However, FLW-Net has fewer parameters (only 17K parameters) and consumes less running time during testing. It should be noted that when testing on the LOL-V2 and LOL-V1 datasets, the hyperparameters of DecNet, KIND and KIND++ are derived from the reference image.

Table 2 presents the color comparison results of various methods on the LOL (V1 (Wei et al., 2018) & V2 (Yang et al., 2021)) datasets. We used the total color difference ΔE_{00} in CIEDE2000 to evaluate different methods. As can be observed, the proposed method achieved lower CIEDE2000(ΔE_{00}) values than other methods,

Table 2 Color comparison results on LOL (V1 (Wei et al., 2018) & V2 (Yang et al., 2021)) datasets.

Method	CIEDE2000↓	
	LOL-V1	LOL-V2
LIME(X. Guo et al., 2016)	14.62	17.63
RetinexNet(Wei et al., 2018)	13.76	18.07
Zerodce++(C. Li et al., 2021)	19.28	14.27
KIND(Y. Zhang et al., 2019)	9.68	6.74
KIND++(Y. Zhang, Guo, et al., 2021)	8.51	9.36
IAT(Cui et al., 2022)	7.97	8.17
DecNet(X. Liu, Xie, et al., 2023)	8.93	8.87
PairLIE(Z. Fu et al., 2023)	11.93	11.23
FLW(Training on LOL V1)	7.74	6.64
FLW(Training on LOL V2)	**7.23**	**6.06**

demonstrating the effectiveness of our proposed method in color restoration.

Table 3 Comparison results on LOL-V2 (Yang et al., 2021)) datasets with fixed hyperparameter in all methods during test. ($\mu_{test} = 0.4$ is adopted in our method)

Method↑	PSNR↑	SSIM	NIQE↓	CIEDE2000↓
KIND(Y. Zhang et al., 2019)	20.59	0.82	4.86	9.3872
KIND++(Y. Zhang, Guo, et al., 2021)	17.66	0.77	4.73	13.89
DecNet(X. Liu, Xie, et al., 2023)	21.13	0.83	4.85	10.45
FLW(Ours)	**23.50**	**0.86**	**3.88**	**8.28**

In Figures 3 and 4, we can observe that the images enhanced by FLW-Net are more closely aligned with the reference images in terms of brightness, contrast, and color. The models used in LIME and RetinexNet are simplified Retinex models. Therefore, the Saturation and Hue of the enhanced images are identical to those of the original low-light images, especially in Figure 4. Although KIND and KIND++ introduced a restoration network to recover the color and remove noise in the reflection image, the results are still not stable. For instance, in Figure 4(e), the image enhanced by KIND++ still has color deviations compared with the reference image. In
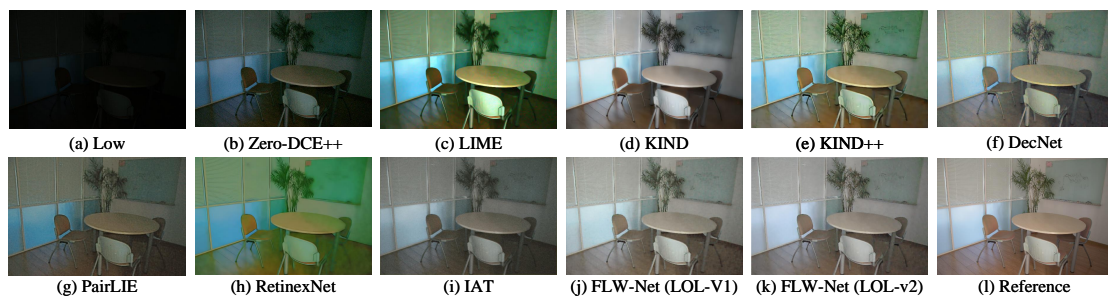
Fig. 4 Visual comparison results on LOL-V2 dataset. Comparison methods include Zero-DCE++, LIME, KIND, KIND++, DecNet, PairLIE, RetienxNet, IAT and the proposed FLW-Net trained on LOL-V1 and LOL-V2 datasets. (Best viewed on high-resolution displays with zoom-in.)

Fig. 5 Visual comparison results on MF dataset (X. Fu et al., 2016). Comparison methods include Zero-DCE++, LIME, KIND, KIND++, DecNet, PairLIE, RetienxNet, IAT and the proposed FLW-Net trained on LOL-V1 and LOL-V2 datasets. (Best viewed on high-resolution displays with zoom-in.)
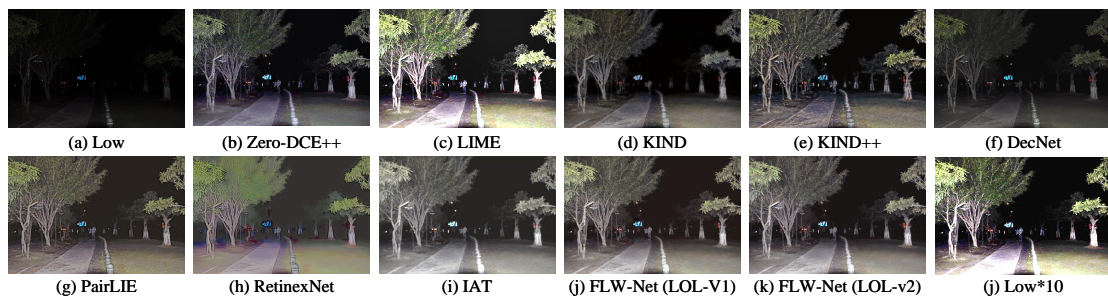

Fig. 6 Visual comparison results on image from Internet. Comparison methods include Zero-DCE++, LIME, KIND, KIND++, DecNet, PairLIE, RetienxNet, IAT and the proposed FLW-Net trained on LOL-V1 and LOL-V2 datasets. (Best viewed on high-resolution displays with zoom-in.)
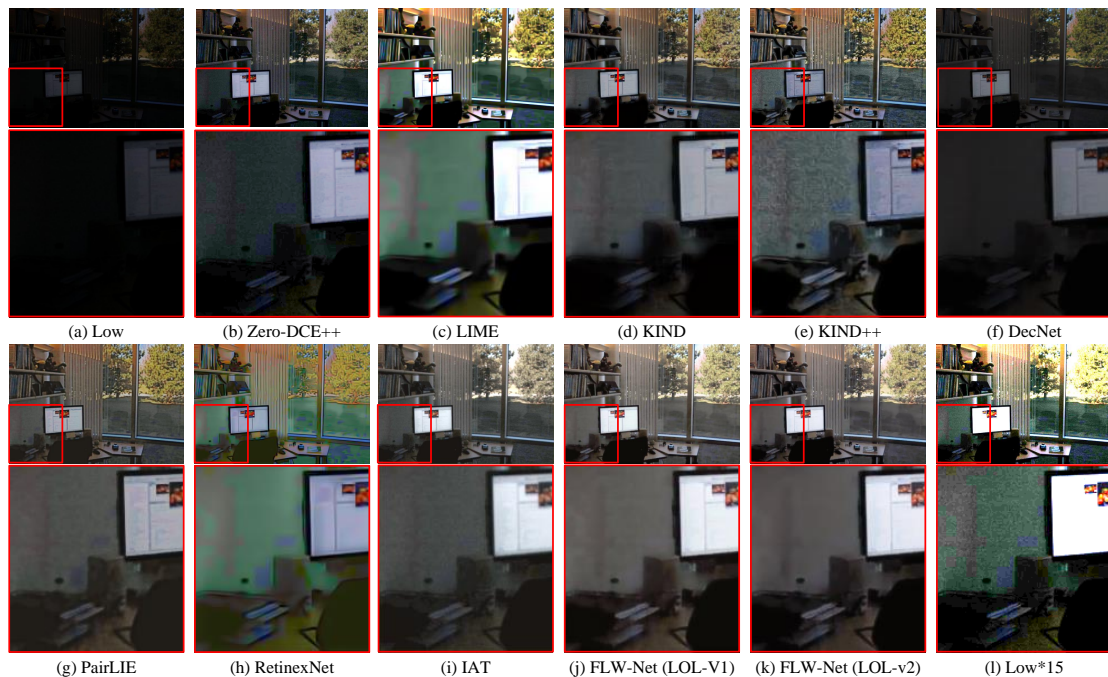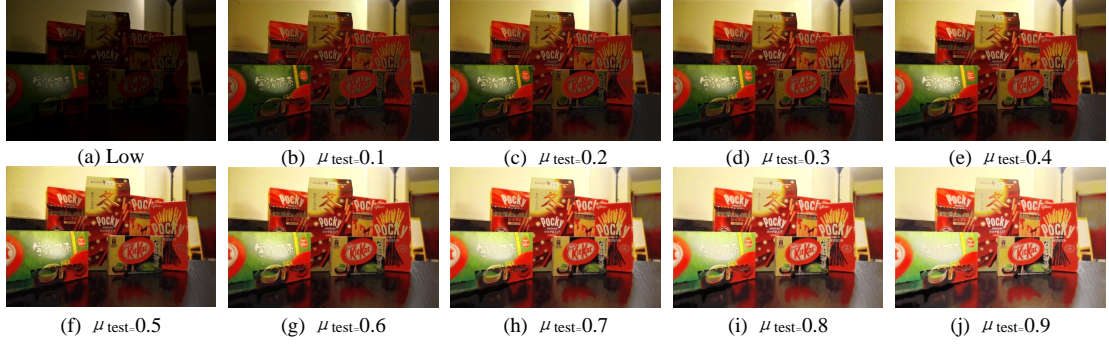
(a) Low   (b) μ_{test}=0.1   (c) μ_{test}=0.2   (d) μ_{test}=0.3   (e) μ_{test}=0.4

(f) μ_{test}=0.5   (g) μ_{test}=0.6   (h) μ_{test}=0.7   (i) μ_{test}=0.8   (j) μ_{test}=0.9

Fig. 7 Visual comparison results on MF dataset when we change the μ_{test} value from 0.1 to 0.9 with a step of 0.1. (Best viewed on high-resolution displays with zoom-in.)

(a) Low   (b) μ_{test}=0.1   (c) μ_{test}=0.2   (d) μ_{test}=0.3   (e) μ_{test}=0.4

(f) μ_{test}=0.5   (g) μ_{test}=0.6   (h) μ_{test}=0.7   (i) μ_{test}=0.8   (j) μ_{test}=0.9
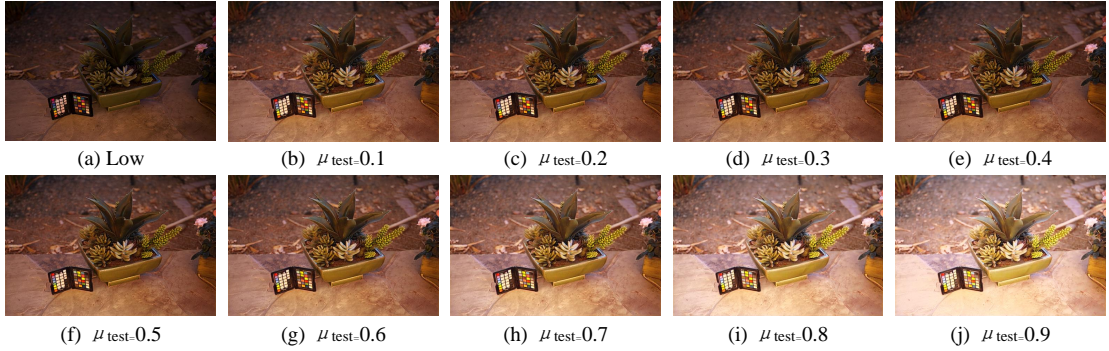
Fig. 8 Visual comparison results on LIME dataset when we change the μ_{test} value from 0.1 to 0.9 with a step of 0.1.(Best viewed on high-resolution displays with zoom-in.)

Figure 5(d), the small light source was treated as noise and removed by KIND.

During testing, the value of μ_{test} can be obtained from reference images such as (Y. Zhang et al., 2019), (X. Liu, Xie, et al., 2023), and (Y. Zhang, Guo, et al., 2021). However, in practical applications, the value of μ_{test} is typically determined by the user or machine to dynamically adjust enhancement results, rather than relying on the reference image. Therefore, it is important to carefully consider the impact of μ_{test} values. In this regard, we also present the results obtained when μ_{test} takes on constant values during testing.

Table 3 shows the quantitative comparison results with other methods which also introduces hyperparameter, and it can be seen that the proposed method can achieve better performance than other methods with less parameters and simpler network structure. Fig. 5 and 6 shows the visual comparison results with other methods. It

can be seen that, the proposed method can effectively remove noise while preserving structural information.

4.3 Ablation Study

We performed two ablation studies on the LOL-V2 dataset to demonstrate the effectiveness of each component in our proposed method. For quantitative comparison, we used the evaluation metrics of PSNR, SSIM and CIEDE2000.

Contribution of Each Loss: In this ablation study, we used the complete network trained with L_1 and SSIM loss as the baseline model. We then incorporated the relative loss functions into the network's loss function and retrained it to examine their effects on the performance of the model. Additionally, we also trained the network solely with the relative loss functions to evaluate their effectiveness.

Table 4 The influence of different training losses. During testing, the input μ_{test} value can be a constant (e.g., $\mu_{test} = 0.4$) or obtained through the reference(e.g., μ_{test} equals the mean value of the reference's V channel).

Loss functions				$\mu_{test} = 0.4$			μ_{test} from reference		
L_1+SSIM	L_{color}	$L_{brightness}$	$L_{structure}$	PSNR↑	SSIM↑	CIEDE2000↓	PSNR↑	SSIM↑	CIEDE2000↓
✓				22.62	0.84	8.67	26.26	0.87	6.20
	✓	✓	✓	20.54	0.83	10.60	21.22	0.84	9.85
✓	✓			22.88	0.85	8.37	26.49	0.87	5.92
✓		✓		**23.72**	0.85	**7.90**	**26.80**	0.87	5.86
✓			✓	23.28	**0.86**	8.14	26.75	**0.88**	**5.84**
✓	✓	✓	✓	23.50	**0.86**	8.28	26.61	**0.88**	6.06

In Table 4, it can be observed that the addition of each relative loss to the baseline model leads to a slight improvement in PSNR or SSIM when μ_{test} is obtained from reference images. However, when μ_{test} takes on a constant value for all testing images, both $L_{brightness}$ and $L_{structure}$ demonstrate significant improvements in PSNR and SSIM. This demonstrates that the proposed two loss functions can help the network better learn noise removal and structural preservation, enhancing the stability of the network. Between these two losses, $L_{brightness}$ shows a better improvement in PSNR, which is related to its denoising ability, while $L_{structure}$ shows a better improvement in SSIM, which is related to its ability to retain structural information. On the other hand, the improvement in PSNR and SSIM with L_{color} is relatively minor. This could be because L_{color} only considers the information of a single pixel, whereas noise removal and retention of structural information require the introduction of information from surrounding pixels.

Furthermore, as shown in Table 4, all three loss functions based on relative information can lead to improvements in color restoration performance. Among them, the proposed $L_{brightness}$ and $L_{structure}$ demonstrate better color restoration performance than the commonly used L_{color}. However, when we use all three loss functions simultaneously, the performance of the CIEDE2000 metric does not reach its optimum. This may be due to conflicts between the assumptions of the models used by the three loss functions. The same issue is also reflected in the PSNR metric. When we only add $L_{brightness}$ to the loss functions, the PSNR value can reach 23.72 and 26.80, while when we add all three loss functions simultaneously, the PSNR value can only reach 23.50 and 26.61.

Fig. 7 and 8 show the impact of different μ_{test} values on the enhancement results. It can be seen that, as μ_{test} value changes, the enhanced images' brightness change accordingly. This indicates that

the strategy of using the parameter μ_{test} to control the enhancement results is feasible. The brightness
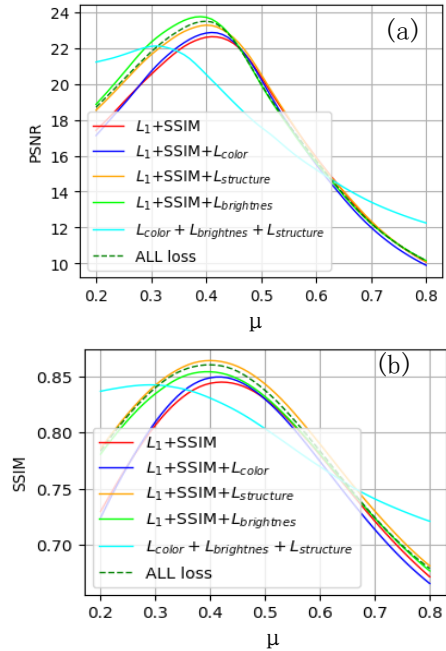
Fig. 9 Influence on PSNR and SSIM of different loss functions when changing the μ_{test} value. (a) The PSNR with different μ_{test} values. (b) The SSIM with different μ_{test} value. ALL loss means that all loss functions are added with the same weight (Equation (12)).

differences have a significant influence on PSNR and SSIM values. Fig. 9 shows the changes in SSIM and PSNR values with different μ_{test} values and loss functions. It can be seen that, when the network is trained without L_1 and SSIM loss, it shows more stable performance in terms of both SSIM and PSNR. If the training losses include L_1 and SSIM loss, the highest values of both SSIM and PSNR are achieved when μ_{test} is close to 0.4.

Table 5 The influence of GFE component and loss functions based on relative information. During training, the input μ_{train} value can be a constant (e.g., $\mu_{train} = 0.4$) or obtained through the reference(e.g., μ_{train} equals the mean value of the reference's V channel in this table). During testing, the input μ_{test} value is a constant for all images in this table ($\mu_{test} = 0.4$). Relative losses represents $L_{color} + L_{brightness} + L_{structure}$

Loss functions		GFE	$\mu_{test} = 0.4$		LOL-V2	
L_1+SSIM	Relative losses	component	$\mu_{train} = 0.4$	μ_{train} from reference	PSNR↑	SSIM↑
✓					18.32	0.80
✓	✓				19.05	0.82
✓		✓	✓		20.59	0.83
✓		✓		✓	22.62	0.84
✓	✓	✓		✓	**23.50**	**0.86**

This is due to the fact that the artificially selected reference images in the LOL-V2 dataset have a mean value of 0.41 in their V channels. There are very few images' V channels with a mean brightness above 0.5, which causes SSIM and PSNR drop significantly when the value of μ_{test} is higher.

Moreover, in Fig. 9(b), we can see that the network trained with L_1, SSIM and $L_{structure}$ loss achieves the highest SSIM. In Fig. 9(a), the network trained with L_1, SSIM and $L_{brightness}$ achieved the highest PSNR. This demonstrates the effectiveness of $L_{structure}$ and $L_{brightness}$ in improving structure and brightness restoration, respectively.

Contribution of GFE component: In this ablation study, the network trained with L_1 and SSIM loss without the GFE component was considered as the baseline model. The effects of adding the GFE component and losses proposed in this paper were then compared and studied. The results are presented in Table 5. It should be noted that the input μ_{test} value is constant for all images during testing in this table.

Table 5 demonstrates that when we add either the other proposed loss functions or the GFE component to the baseline model, both PSNR and SSIM values show improvement. This provides strong evidence for the effectiveness of the GFE component and the loss functions designed with relative information. The GFE component can capture global brightness information and integrate it into the enhancement process with few parameters, leading to a significant improvement in PSNR by 2.27 dB and SSIM by 0.03 (PSNR: $18.32 \rightarrow 20.59$, SSIM: $0.80 \rightarrow 0.83$), even with fixed μ_{test} and μ_{train} values during testing and training.

Furthermore, we observed that when the μ_{train} value is calculated from the reference images

during training, the improvement in PSNR and SSIM values is even more significant (e.g., PSNR: $18.32 \rightarrow 22.63$, SSIM: $0.80 \rightarrow 0.84$). This finding highlights the challenge posed by the one-to-many problem in learning how to remove noise and retain structural information in low-light image enhancement. Therefore, the GFE component, which is a simple and efficient method for connecting the input and output images, can greatly improve the enhancement results by facilitating the learning process.

4.4 Combined With Other Networks

The proposed $L_{brightness}$ and $L_{structure}$ loss functions can also be combined with other supervised image enhancement methods to improve their performance with fewer operations or parameters . For example, BM3D or additional sub-network is adopted to denoise on the reflectance image in Retinex based low-light image enhancement methods(e.g., RetinexNet (Wei et al., 2018), KIND (Y. Zhang et al., 2019), and KIND++(C. Li et al., 2021)). However, if we add the proposed two relative loss functions into the training of the network, their performace can be imporved without the additional part. Table 6, 7 and Fig. 10, 11 show the examples of combining $L_{brightness}$ and $L_{structure}$ with RetinexNet and KIND.

RetinexNet was trained with LOL-V1 dataset in the original paper (Wei et al., 2018), and BM3D was used for denoising reflectance images. By adding the two proposed relative loss functions to the training of RetinexNet, the network achieves better PSNR, SSIM, and CIEDE2000 scores than the original RetinexNet and RetinexNet combined with BM3D, as Table 6 shows. Also, as seen in Fig. 10, RetinexNet with $L_{brightness}$ and $L_{structure}$ loss functions can significantly reduce noise in the enhanced image (Fig. 10(b) and (d)). Compared

Table 6 The influence of combining the $L_{brightness}$ and $L_{structure}$ with RetinexNet on LOL-V1 dataset. In the original paper(Wei et al., 2018), RetienxNet was trained with LOL-V1 dataset, and the authors used BM3D to denoise the reflectance image.

Original loss	BM3D	$L_{brightness}$&$L_{structure}$	PSNR↑	SSIM↑	CIEDE2000↓
✓			16.79	0.42	15.89
✓	✓		17.86	0.78	13.76
✓		✓	**18.87**	**0.79**	**12.01**

(a) Low   (b) RetinexNet w/o BM3D   (c) RetinexNet & BM3D   (d) RetinexNet & Our loss   (e) Reference

Fig. 10 The visual effect of combining the proposed loss functions with RetinexNet. (a) The input low-light images. (b) Results of RetinexNet without BM3D. (c) Results of RetinexNet with BM3D. (d) Results of RetinexNet without BM3D and with $L_{brightness}$ and $L_{structure}$ in the training loss function. (e) Reference images. In the original paper (Wei et al., 2018), RetinexNet was trained with LOL-V1 dataset and BM3D was used to denoise on the reflectance image. (Best viewed on high-resolution displays with zoom-in.)

with the BM3D method (Fig. 10(c)), proposed relative loss functions can help preserve more details and restore the color (Fig. 10(d)).

KIND consists of three sub-networks: Decom-Net, RestorNet, and AdjustNet. By adding the proposed relative loss functions to the training of DecomNet, DecomNet achieved better results than the entire KIND method, as shown in Table 7. At that time, the number of parameters can be reduced by nearly 95%. Figure 11 show some images of combining the relative loss functions with DecomNet of KIND. It can be seen that, the relative loss functions have positive impact on noise reduction, color correction, and detail preservation. Thus, we can obtain comparable or superior results using a simpler network structure with the two proposed relative loss functions.

5 Conclusion

In this paper, we have demonstrated that by using efficient global feature extraction and the proposed relative loss functions, a simple network structure can be employed to achieve image enhancement that is comparable to, or even better than, the current state-of-the-art (SOTA) methods with much less running time. Moreover, the Global Feature Extraction component and loss functions can be combined with other low-light image enhancement techniques to enhance objective evaluation indicators such as PSNR and SSIM. The experimental results show the effectiveness and advantages of our method for low-light image enhancement. However, our approach still has some limitations, such as the dependence of the final enhancement result on the desired brightness parameter μ_{test} and the requirement for paired data during training. To address the first issue, parameters can be automatically selected, as demonstrated in our previous work (Q. Fu et al., 2020), or the contrast of the enhanced image can be further adjusted through GAMMA Correction or other local tone mapping techniques (Zeng et al., 2020). In future research, we aim to

Table 7 The influence of combining the $L_{brightness}$ and $L_{structure}$ with KIND on LOL-V1 dataset. In the original paper(Y. Zhang et al., 2019), KIND was trained with LOL-V1 dataset. KIND has three sub-networks, which are the Decomposition Network, the Restoration Network, and the Illumination Adjustment Network. They are abbreviated as DecomNet, RestorNet, and AdjustNet in this table.

DecomNet	RestorNet&AdjustNet	$L_{brightness}$&$L_{structure}$	PSNR↑	SSIM↑	CIEDE2000↓	Params(M)↓
✓			15.68	0.50	18.73	0.43
✓	✓		17.65	0.78	12.49	8.21
✓		✓	**18.94**	**0.80**	**11.93**	0.43

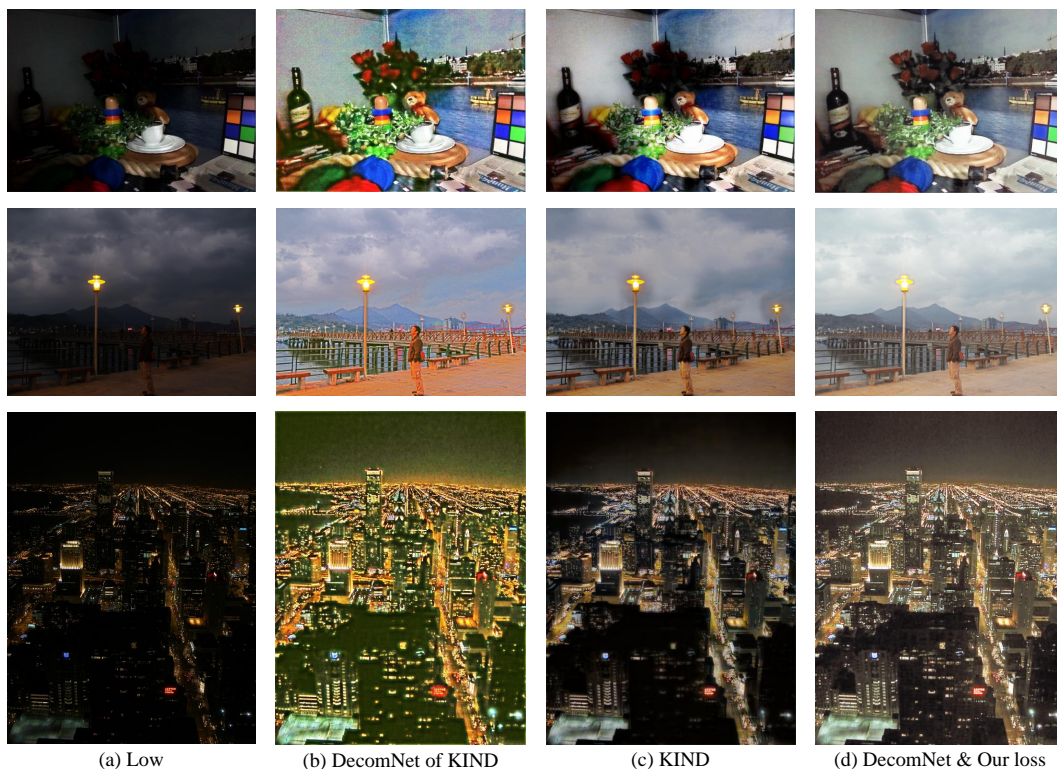(a) Low   (b) DecomNet of KIND   (c) KIND   (d) DecomNet & Our loss

Fig. 11 The visual effect of combining the loss function with DecomNet of KIND. KIND has three sub-networks, which are the Decomposition Network, the Restoration Network, and the Illumination Adjustment Network. They are abbreviated as DecomNet, RestorNet, and AdjustNet in this table. (a) The input low-light images. (b) Results of DecomNet in KIND. (c) KIND. (d) DecomNet in KIND with $L_{brightness}$ and $L_{structure}$ in the original loss function. (Best viewed on high-resolution displays with zoom-in.)

improve the robustness of the proposed method and explore unsupervised approaches for network training.

Declarations

- The authors have no competing interests to declare that are relevant to the content of this article.
- The code is available at https://github.com/hitzhangyu/FLW-Net
- The data used in the paper is sourced from publicly published papers or publicly available websites. However, if there are reasonable justifications, it is also possible to request the data from the corresponding authors.

chinaXiv:202310.03347v1

- Y.Zhang designed the workflow, and wrote the manuscript. J.Wu., Y.Wang., Y.Li., and R.Fu. provided some ideas and solved some problems in the work, and contributed to the manuscript. Y.Xu. and Y.Guo provide some ideas and contributed to the manuscript. X.Xiao, and C.Wang. supervised the work and contributed to the manuscript.

References

Barnard, K., Cardei, V., Funt, B. (2002). A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data. *IEEE Transactions on Image Processing*, *11*(9), 972-984.

https://doi.org/10.1109/TIP.2002.802531

Celik, T., & Tjahjadi, T. (2011). Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, *20*(12), 3431–3441.

Chen, C., Chen, Q., Xu, J., Koltun, V. (2018). Learning to see in the dark. *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 3291–3300).

Cui, Z., Li, K., Gu, L., Su, S., Gao, P., Jiang, Z., . . . Harada, T. (2022). You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. *33rd british machine vision conference 2022, BMVC 2022, london, uk, november 21-24, 2022.* BMVA Press. Retrieved from https://bmvc2022.mpi-inf.mpg.de/0238.pdf

Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K. (2007). Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, *16*(8), 2080–2095.

Dong, X., Wang, G., Pang, Y., Li, W., Wen, J., Meng, W., Lu, Y. (2011). Fast efficient algorithm for enhancement of low lighting video. *2011 ieee international conference on multimedia and expo* (pp. 1–6).

Fu, G., Duan, L., Xiao, C. (2019). A hybrid l2-lp variational model for single low-light image enhancement with bright channel prior. *2019 ieee international conference on image processing (icip)* (pp. 1925–1929).

Fu, Q., Di, X., Zhang, Y. (2020). Learning an adaptive model for extreme low-light raw image processing. *IET Image Processing*, *14*(14), 3433–3443.

Fu, X., Zeng, D., Huang, Y., Liao, Y., Ding, X., Paisley, J. (2016). A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, *129*, 82–96.

Fu, Z., Yang, Y., Tu, X., Huang, Y., Ding, X., Ma, K.-K. (2023). Learning a simple low-light image enhancer from paired low-light instances. *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 22252–22261).

Gijsenij, A., Gevers, T., van de Weijer, J. (2011). Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing*, *20*(9), 2475-2489.

https://doi.org/10.1109/TIP.2011.2118224

Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 1780–1789).

Guo, X., Li, Y., Ling, H. (2016). Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, *26*(2), 982–993.

He, J., Liu, Y., Qiao, Y., Dong, C. (2020). Conditional sequential modulation for efficient

global image retouching. *European conference on computer vision* (pp. 679–695).

Jiang, N., Lin, J., Zhang, T., Zheng, H., Zhao, T. (2023). Low-light image enhancement via stage-transformer-guided network. *IEEE Transactions on Circuits and Systems for Video Technology*, 1-1.

https://doi.org/10.1109/TCSVT.2023.3239511

Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., . . . Wang, Z. (2021). Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, *30*, 2340–2349.

Kimmel, R., Elad, M., Shaked, D., Keshet, R., Sobel, I. (2003). A variational framework for retinex. *International Journal of computer vision*, *52*(1), 7–23.

Kumar, R., & Bhandari, A.K. (2022). Fuzzified contrast enhancement for nearly invisible images. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(5), 2802-2813.

https://doi.org/10.1109/TCSVT.2021.3098763

Lee, C., Lee, C., Kim, C.-S. (2013). Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, *22*(12), 5372–5384.

Li, C., Guo, C., Loy, C.C. (2021). Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(8), 4225–4238.

Li, M., Liu, J., Yang, W., Sun, X., Guo, Z. (2018). Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*,

27(6), 2828–2841.

Liu, C., Wu, F., Wang, X. (2022). Efinet: Restoration for low-light images via enhancement-fusion iterative network. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(12), 8486-8499.

https://doi.org/10.1109/TCSVT.2022.3195996

Liu, X., Ma, W., Ma, X., Wang, J. (2023). Laenet: A locally-adaptive embedding network for low-light image enhancement. *Pattern Recognition*, *133*, 109039.

https://doi.org/10.1016/j.patcog.2022.109039

Liu, X., Xie, Q., Zhao, Q., Wang, H., Meng, D. (2023). Low-light image enhancement by retinex-based algorithm unrolling and adjustment. *IEEE Transactions on Neural Networks and Learning Systems*, 1-14.

https://doi.org/10.1109/TNNLS.2023.3289626

Liu, Y., He, J., Chen, X., Zhang, Z., Zhao, H., Dong, C., Qiao, Y. (2022). Very lightweight photo retouching network with conditional sequential modulation. *IEEE Transactions on Multimedia*.

Luo, M.R., Cui, G., Rigg, B. (2001). The development of the cie 2000 colour-difference formula: Ciede2000. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, *26*(5), 340–350.

Ma, L., Ma, T., Liu, R., Fan, X., Luo, Z. (2022). Toward fast, flexible, and robust low-light image enhancement. *Proceedings of the*

ieee/cvf conference on computer vision and pattern recognition (pp. 5637–5646).

Mittal, A., Soundararajan, R., Bovik, A.C. (2012). Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, *20*(3), 209–212.

Park, S., Yu, S., Moon, B., Ko, S., Paik, J. (2017). Low-light image enhancement using variational optimization-based retinex model. *IEEE Transactions on Consumer Electronics*, *63*(2), 178–184.

Sharma, G., Wu, W., Dalal, E.N. (2005). The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, *30*(1), 21–30.

Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., Li, Y. (2022). Maxim: Multi-axis mlp for image processing. *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 5769–5780).

Wang, S., Zheng, J., Hu, H.-M., Li, B. (2013). Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, *22*(9), 3538–3548.

Wang, Y., Li, X., Xu, K., He, D., Zhang, Q., Li, F., Ding, E. (2022). Neural color operators for sequential image retouching. *arXiv preprint arXiv:2207.08080*.

Wang, Y., Wan, R., Yang, W., Li, H., Chau, L.-P., Kot, A. (2022, Jun.). Low-light image

enhancement with normalizing flow. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(3), 2604-2612.

https://doi.org/10.1609/aaai.v36i3.20162

Wei, C., Wang, W., Yang, W., Liu, J. (2018). Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*.

Xiong, W., Liu, D., Shen, X., Fang, C., Luo, J. (2020). Unsupervised real-world low-light image enhancement with decoupled networks. *arXiv preprint arXiv:2005.02818*.

Xu, K., Chen, H., Xu, C., Jin, Y., Zhu, C. (2022). Structure-texture aware network for low-light image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(8), 4983-4996.

https://doi.org/10.1109/TCSVT.2022.3141578

Yang, W., Wang, W., Huang, H., Wang, S., Liu, J. (2021). Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, *30*, 2072–2086.

Yu, S.-Y., & Zhu, H. (2019). Low-illumination image enhancement algorithm based on a physical lighting model. *IEEE Transactions on Circuits and Systems for Video Technology*, *29*(1), 28-37.

https://doi.org/10.1109/TCSVT.2017.2763180

Zeng, H., Cai, J., Li, L., Cao, Z., Zhang, L. (2020). Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhang, K., Yuan, C., Li, J., Gao, X., Li, M. (2023). Multi-branch and progressive network for

chinaXiv:202310.03347v1

19

low-light image enhancement. *IEEE Transactions on Image Processing*, *32*, 2295-2308.

https://doi.org/10.1109/TIP.2023.3266171

Zhang, Y., Di, X., Zhang, B., Ji, R., Wang, C. (2022). Better than reference in low-light image enhancement: Conditional re-enhancement network. *IEEE Transactions on Image Processing*, *31*, 759-772.

https://doi.org/10.1109/TIP.2021.3135473

Zhang, Y., Di, X., Zhang, B., Li, Q., Yan, S., Wang, C. (2021). Self-supervised low light image enhancement and denoising. *arXiv preprint arXiv:2103.00832*.

Zhang, Y., Guo, X., Ma, J., Liu, W., Zhang, J. (2021). Beyond brightening low-light images. *International Journal of Computer Vision*, *129*(4), 1013–1037.

Zhang, Y., Zhang, J., Guo, X. (2019). Kindling the darkness: A practical low-light image enhancer. *Proceedings of the 27th acm international conference on multimedia* (pp. 1632–1640).

Zhou, Z., Shi, Z., Ren, W. (2023). Linear contrast enhancement network for low-illumination image enhancement. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1-16.

https://doi.org/10.1109/TIM.2022.3232641